

Foundations of Artificial Intelligence

Prof. Dr. W. Burgard, Prof. Dr. M. Riedmiller
B. Frank, D. Joho, Dr. S. Lange
Summer Term 2011

University of Freiburg
Department of Computer Science

Exercise Sheet 5

Due: Friday, July 8, 2011 (in the exercise session)

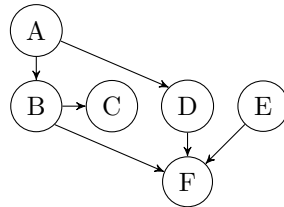
Exercise 5.1 (Conditional independence)

This exercise investigates the way in which conditional independence relationships affect the amount of information needed for probabilistic calculations.

- (a) Suppose we wish to calculate $\mathbf{P}(X|E_1, E_2)$ and we have no conditional independence information. Which of the following sets of numbers are sufficient for the calculation?
- (i) $\mathbf{P}(E_1, E_2), \mathbf{P}(X), \mathbf{P}(E_1|X), \mathbf{P}(E_2|X)$
 - (ii) $\mathbf{P}(E_1, E_2), \mathbf{P}(X), \mathbf{P}(E_1, E_2|X)$
 - (iii) $\mathbf{P}(X), \mathbf{P}(E_1|X), \mathbf{P}(E_2|X)$
- (b) Suppose we know that $\mathbf{P}(E_1|X, E_2) = \mathbf{P}(E_1|X)$ for all values of X, E_1 , and E_2 . Now which of the three sets are sufficient?

Exercise 5.2 (Bayesian Networks)

Consider the following Bayesian network:



- (a) Rewrite the joint probability distribution $P(A, B, C, D, E, F)$ using the conditional independencies expressed by the network.
- (b) Suppose that all the random variables A, B, C, D, E, F in the Bayesian network can only have two possible values *yes* and *no*. What's the minimum number of probabilities required to fully define the Bayesian network whose structure is given above?
- Hint:* Remember that e.g. $P(E = \textit{yes}) = 1 - P(E = \textit{no})$.
- (c) How many probabilities would be required to define the full joint distribution over A, B, C, D, E, F if we could not assume the conditional independencies expressed by the Bayesian network?

Exercise 5.3 (Decision trees)

(a) Specify decision trees representing the following Boolean functions:

(1) $A \wedge \neg B$

(2) $A \vee (B \wedge C)$

(3) $A \text{ XOR } B$

(4) $(A \wedge B) \vee (C \wedge D)$

Exercise 5.4 (Attribute selection)

Here we will practice the basic information-theoretical concepts used to build decision trees. Consider the following set of training examples

a_1	a_2	Classification
T	T	+
T	T	+
T	F	-
F	F	+
F	T	-
F	T	-

What is the information gain of a_1 and a_2 ? Draw the complete decision tree for this data set and use the information gain to select the attribute.

The exercise sheets may and should be handed in and be worked on in groups of three (3) students. Please fill the cover sheet¹ and attach it to your solution.

¹<http://ais.informatik.uni-freiburg.de/teaching/ss11/ki/cover-sheet.pdf>